NATIONAL BUREAU OF STANDARDS REPORT

2751

STOCHASTIC SEARCH FOR THE MAXIMUM OF A FUNCTION

by

E. W. Barankin

U. S. DEPARTMENT OF COMMERCE
NATIONAL BUREAU OF STANDARDS

U. S. DEPARTMENT OF COMMERCE

Sinclair Weeks, *Secretary*

NATIONAL BUREAU OF STANDARDS

A. V. Astin, *Director*

THE NATIONAL BUREAU OF STANDARDS

The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work. In general, each section is engaged in specialized research, development, and engineering in the field indicated by its title. A brief description of the activities, and of the resultant reports and publications, appears on the inside of the back cover of this report.

Electricity. Resistance Measurements. Inductance and Capacitance. Electrical Instruments. Magnetic Measurements. Applied Electricity. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Gage.

Heat and Power. Temperature Measurements. Thermodynamics. Cryogenics. Engines and Lubrication. Engine Fuels. Cryogenic Engineering.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Measurements. Infrared Spectroscopy. Nuclear Physics. Radioactivity. X-Rays. Betatron. Nucleonic Instrumentation. Radiological Equipment. Atomic Energy Commission Instruments Branch.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Gas Chemistry. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Aerodynamics. Engineering Mechanics. Hydraulics. Mass. Capacity, Density, and Fluid Meters.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Organic Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion.

Mineral Products. Porcelain and Pottery. Glass. Refractories. Enameled Metals. Concreting Materials. Constitution and Microstructure. Chemistry of Mineral Products.

Building Technology. Structural Engineering. Fire Protection. Heating and Air Conditioning. Floor, Roof, and Wall Coverings. Codes and Specifications.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Machine Development.

Electronics. Engineering Electronics. Electron Tubes. Electronic Computers. Electronic Instrumentation.

Radio Propagation. Upper Atmosphere Research. Ionospheric Research. Regular Propagation Services. Frequency Utilization Research. Tropospheric Propagation Research. High Frequency Standards. Microwave Standards.

Ordnance Development.
Electromechanical Ordnance.
Ordnance Electronics.
These three divisions are engaged in a broad program of research and development in advanced ordnance. Activities include basic and applied research, engineering, pilot production, field testing, and evaluation of a wide variety of ordnance matériel. Special skills and facilities of other NBS divisions also contribute to this program. The activity is sponsored by the Department of Defense.

Missile Development. Missile research and development: engineering, dynamics, intelligence, instrumentation, evaluation. Combustion in jet engines. These activities are sponsored by the Department of Defense.

- Office of Basic Instrumentation
- Office of Weights and Measures.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT **NBS REPORT**

1101-10-5102 August 12, 1953 2751

STOCHASTIC SEARCH FOR THE MAXIMUM OF A FUNCTION*

by

E. W. Barankin

University of California, Berkeley, National Bureau of Standards, Los Angeles, and Logistics Research Project, University of California, Los Angeles

PREPRINT

*The preparation of this paper was sponsored (in part) by ARDC

STOCHASTIC SEARCH FOR THE MAXIMUM OF A FUNCTION[1]

by

E. W. Barankin[2]

1. INTRODUCTION

Consider a real valued function f on a domain D. We assume that D is a finite set of points, say N in number. Let $M = \max_{x \in D} f(x)$, and suppose f takes on the value M at exactly ν points; the set of these points will be denoted by K.

Far from looking upon our assumption of finiteness of D as a specialization we consider that it reflects the real computing situation. Consequently, our results pertain also to situations in which D is taken to be a continuum and in which continuous distributions are employed; it is required only to recognize the discontinuous distributions to which the given continuous distributions are approximations.

It is our purpose here to study stochastic processes $\mathcal{S} = \left\{X_i, \ i = 1, 2, \cdots\right\}$, where each X_i is a random variable in D; more particularly, to study the implications of such a process for the determination of M when the realized values x_i of the random variables X_i are submitted to computation of f- values. To be precise, we suppose that after each stage of the process, say the i-th, if the point $x_i \in D$ is the realized value of X_i, we compute $f(x_i)$. Thus,

---

[1]The preparation of this paper was sponsored (in part) by ARDC.

[2]University of California, Berkeley, National Bureau of Standards, Los Angeles, and Logistics Research Project, University of California, Los Angeles.

after the k-th stage, if the realized values of $X_1, X_2, \cdots, X_k$ are $x_1, x_2, \cdots, x_k$, respectively, then we have before us the f- values $f(x_1), f(x_2), \cdots, f(x_k)$. This represents a certain amount of information about the function f; let us denote it by $\mathcal{J}_k^{\,o}(x_1, x_2, \cdots, x_k)$.

It may be that some general knowledge concerning f can be coupled with the information $\mathcal{J}_k^{\,o}(x_1, x_2, \cdots, x_k)$ to give still a greater amount of specific useful information after the k-th stage. Accordingly, we shall take the k-th stage information scheme to be in general as follows, in addition to $\mathcal{J}_k^{\,o}$:

$\mathcal{J}_k^{\,1}(x_1, x_2, \cdots, x_k)$: a known number $T_k(x_1, x_2, \cdots, x_k) \geq \max\limits_{i=1,2,\cdots,k} f(x_i)$ such that it is known that $M \geq T_k(x_1, x_2, \cdots, x_k)$;

$\mathcal{J}_k^{\,2}(x_1, x_2, \cdots, x_k)$: a known subset of D, $A_{k+1}(x_1, x_2, \cdots, x_k) \supseteq \{x_1, x_2, \cdots, x_k\}$ such that for each $x \in A_{k+1}(x_1, x_2, \cdots, x_k)$, although the number $f(x)$ may not be known exactly, it is known that $f(x) \leq T_k(x_1, x_2, \cdots, x_k)$;

$\mathcal{J}_k^{\,3}(x_1, x_2, \cdots, x_k)$: a known subset of D, $B_{k+1}(x_1, x_2, \cdots, x_k)$, disjoint from $A_{k+1}(x_1, x_2, \cdots, x_k)$, such that for each $x \in B_{k+1}(x_1, x_2, \cdots, x_k)$ the number $f(x)$ is not known exactly,but it is known that $f(x) > T_k(x_1, x_2, \cdots, x_k)$.

The special case of this scheme, wherein

$$T_k(x_1, x_2, \cdots, x_k) = \max_{i=1,2,\cdots,k} f(x_i) \quad ,$$

$$A_{k+1}(x_1, x_2, \cdots, x_k) = \left\{ x_1, x_2, \cdots, x_k \right\} \quad ,$$

and

$$B_{k+1}(x_1, x_2, \cdots, x_k) = \Lambda$$

(where Λ denotes the empty subset of D) is that in which $\mathcal{I}_k{}^o(x_1, x_2, \cdots, x_k)$ is actually the total information at hand.

On the basis of the above information scheme, an immediately suggested estimated value for M is $T_k(x_1, x_2, \cdots, x_k)$. There are alternatives to this; for example, the maximum of the function on D obtained by a specified extrapolation of the known values of f on $A_{k+1}(x_1, x_2, \cdots, x_k)$. However, in this article we shall fix our attention on the estimator $\left\{ T_i, \; i = 1, 2, \cdots \right\}$. We are consequently committed to study the stochastic convergence of the sequence of random variables $\mathcal{T} = \left\{ T_1(X_1), T_2(X_1, X_2), \cdots \right\}$.

Of course, we have for all $k = 1, 2, \cdots$, and all $x_1, x_2, \cdots, x_k, x_{k+1}$,

(1.1) $$T_k(x_1, x_2, \cdots, x_k) \leq M$$

and

(1.2) $$T_{k+1}(x_1, x_2, \cdots, x_{k+1}) \geq T_k(x_1, x_2, \cdots, x_k) \quad .$$

From (1.2) it follows that

$$(1.3)\qquad A_{k+2}(x_1, x_2, \cdots, x_{k+1}) \supseteq A_{k+1}(x_1, x_2, \cdots, x_k) \ .$$

The stochastic process $\mathcal{S}$ that will be at work producing points of D, will be conditioned by the computer's behavior. This conditioning is effected upon the conditional probabilities

$$(1.4)\qquad P\left\{X_{k+1} = x \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} \ , x \in D \ .$$

The conditioning is realized through the choice, by the computer, of random devices which (he guesses) will secure desirable analytic conditions on the conditional probabilities (1.4). Just what such desirable analytic conditions are will be determined by studying the implications of various analytic conditions on (1.4) in the light of our designated criteria for the process $\mathcal{S}$. Our criteria in this article are the following: large values for the probabilities

$$(1.5)\qquad P\left\{T_k(X_1, X_2, \cdots, X_k) = M\right\} \ , k = 1, 2, \cdots ,$$

and weak and strong stochastic convergences of $\mathcal{T}$ to M. Thus, the problem that will occupy us is this: <u>to find analytic conditions on (1.4) which will give large values to the numbers (1.5) and which will insure weak and strong stochastic convergences of $\mathcal{T}$ to M.</u>

Now, there is, of course, a limitation on the desirable analytic conditions for which we may seek, a limitation that must be observed if the computer is to be able to realize these conditions in the choices (of random devices) that are open to him; namely, the fulfillment of

the conditions must be feasible under the information at hand at each stage. To illustrate this, consider that after any stage, say the k-th, it would be highly desirable to select random devices to secure

$$P\left\{X_{k+1} \varepsilon K \mid X_i = x_i, \quad i = 1, 2, \cdots, k\right\} = 1 \tag{1.6}$$

whatever be $x_1, x_2, \cdots, x_k$. But unless

$$\begin{cases} A_{k+1}(x_1, x_2, \cdots, x_k) = D \\ \text{or} \\ \begin{cases} B_{k+1}(x_1, x_2, \cdots, x_k) = \bar{A}_{k+1}(x_1, x_2, \cdots, x_k) \text{ and it} \\ \text{is known that } f \text{ is constant on } B_{k+1}(x_1, x_2, \cdots, x_k) \end{cases} \end{cases} \tag{1.7}$$

(where $\bar{A}_{k+1}$ denotes the complement of A_{k+1} in D) we do not have enough information to locate a subset of K, therefore not enough information to secure (1.6). It will be seen in our discussion that our conditions are feasible in the sense here indicated.

Our results are stated in Theorem 1 (end of section 2) and in Theorem 2 (end of section 3). And in section 4 we discuss the matter of improving these results.

2. SUFFICIENT CONDITIONS FOR STRONG CONVERGENCE OF $\mathcal{J}$ TO M

As of each k-th stage, the computer is motivated in two ways toward putting feasible analytic conditions on the probabilities (1.4). First, if neither of the situations (1.7) holds, he wishes to insure a positive (k+1)-th stage probability to each point of

$\bar{A}_{k+1}(x_1, x_2, \cdots, x_k)$; this, by way of not ruling out the possibility of any such point being realized in the (k+1)-th stage, since (according to the information at hand) f may have the value M only at one of these complementary points. Secondly, he wishes to assign high probabilities to sets concerning which he has been led to believe that in them the values of f are higher than the f- values he has already discovered. The conditions we shall put on the probabilities (1.4) will embody these motivations.

But first, consider (1.7). If either of these situations holds, and only if one of them holds, we know exactly which points constitute K. And it is then feasible to make the probability of K unity. Thus, we shall take

$$(2.1) \quad P\left\{X_{k+1} \varepsilon\ K \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} = 1 \ \underline{\text{if}}\ (1.7)\ \underline{\text{holds}} \quad .$$

Now consider the first motivation. One feasible realization of this is in terms of a chosen positive number $c_{k+1} < 1$, and is as follows: Whatever be $x_1, x_2, \cdots, x_k$, if (1.7) does not hold, we take

$$(2.2) \quad P\left\{X_{k+1} = x \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} \geqq c_{k+1} \ \underline{\text{for}}\ \underline{\text{each}} \quad x\ \varepsilon\ \bar{A}_{k+1}(x_1, x_2, \cdots, x_k) \quad .$$

The nature of this condition is such as to imply a sharper restriction on c_{k+1}. Let $\rho(x_1, x_2, \cdots, x_k)$ denote the number of points in $A_{k+1}(x_1, x_2, \cdots, x_k)$. (This number is known under the information

at hand). Then, according to (2.2), the conditional probability of the set $\bar{A}_{k+1}(x_1, x_2, \cdots, x_k)$ is greater than or equal to

$$[N - \rho(x_1, x_2, \cdots, x_k)]\, c_{k+1} ,$$

and therefore this number must be less than or equal to unity, whence

$$c_{k+1} \leq \frac{1}{N - \rho(x_1, x_2, \cdots, x_k)} .$$

Moreover, since c_{k+1} is taken to be independent of $x_1, x_2, \cdots, x_k$, we must have

$$(2.3) \qquad c_{k+1} \leq \frac{1}{N - \min\limits_{x_1, x_2, \cdots x_k} \rho(x_1, x_2, \cdots, x_k)} .$$

We cannot say anything more explicit about the number $\min\limits_{x_1, x_2, \cdots, x_k} \rho(x_1, x_2, \cdots, x_k)$ without further assumptions concerning the sets $A_{k+1}(x_1, x_2, \cdots, x_k)$. As an example, if

$$A_{k+1}(x_1, x_2, \cdots, x_k) = \{x_1, x_2, \cdots, x_k\}$$

for all $x_1, x_2, \cdots, x_k$, then $\min\limits_{x_1, x_2, \cdots, x_k} \rho(x_1, x_2, \cdots, x_k) = 1$, and we get

$$(2.4) \qquad c_{k+1} \leq \frac{1}{N - 1} .$$

Of course, the right-hand side of (2.4) is not greater than the right-hand side of (2.3); consequently, if we required the fulfillment of (2.4), then (2.3) would be automatically satisfied.

We now go to the second motivation. And this leads us immediately to favor at least the set $B_{k+1}(x_1, x_2, \cdots, x_k)$. That is, if this set is not empty, then we know that the f- values at the points in it are greater than the largest value of f so far discovered, and therefore we want to give the points of B_{k+1} a "better than minimal chance" of being realized. If the second of the two situations (1.7) holds, then we know that $B_{k+1}(x_1, x_2, \cdots, x_k) = K$, and we proceed according to (2.1). But if (1.7) does not hold, we do not wish to throw all the conditional probability into B_{k+1}, for, according to our incomplete information, the value M may very well be taken on only in $\overline{A_{k+1} \cup B_{k+1}}$. Thus, we shall want only this much: to give each point of B_{k+1} a probability $\geq$ a positive number, say b_{k+1}, *this number being greater than* c_{k+1}:

$$(2.5) \qquad b_{k+1} > c_{k+1} \quad , \quad k = 1, 2, \cdots \quad .$$

Thus do we favor the points of B_{k+1} in particular, over the points of $\bar{A}_{k+1}$ in general.

But--we repeat--K may not intersect B_{k+1} at all. The computer may have some feeling about this, not at all based on the formal information at hand. This is to say that he may feel justified in making a guess over and above any logical justification by his information. We want to allow for such guessing, and we do so as follows: *there shall be designated a set* $F_{k+1}(x_1, x_2, \cdots, x_k)$, *disjoint from* B_{k+1}, *such that each of its points, as well as the points of* B_{k+1}, *shall be given a conditional probability* $\geq b_{k+1}$.

We have now carried out fully the second motivation, and it has led to the following formal prescription: <u>whatever be</u> $x_1, x_2, \cdots, x_k$, <u>if</u> (1.7) <u>does not hold, we take</u>

(2.6)
$$P\{X_{k+1} = x \mid X_i = x_i, 1, 2, \cdots, k\} \geq b_{k+1} \text{ for each } x \in G_{k+1}(x_1, x_2, \cdots, x_k)$$

<u>where</u>

$$G_{k+1}(x_1, x_2, \cdots, x_k) = B_{k+1}(x_1, x_2, \cdots, x_k) \cup F_{k+1}(x_1, x_2, \cdots, x_k) .$$

If $\tau(x_1, x_2, \cdots, x_k)$ denotes the number of points in the set $G_{k+1}(x_1, x_2, \cdots, x_k)$ then we must have

(2.7)
$$b_{k+1} \leq \frac{1}{\max\limits_{x_1, x_2, \cdots, x_k} \tau(x_1, x_2, \cdots, x_k)} .$$

It is to be remarked that if B_{k+1} is empty then F_{k+1} might very well include the set

(2.8)
$$A^*_{k+1}(x_1, x_2, \cdots, x_k) = \left\{ x \in A_{k+1}(x_1, x_2, \cdots, x_k) \;\middle|\; \begin{array}{l} x \text{ is known to have the property that} \\ f(x) = T_k(x_1, x_2, \cdots, x_1), \\ \text{or} \\ x \text{ is not known to have the property that} \\ f(x) < T_k(x_1, x_2, \cdots, x_k) \end{array} \right\} .$$

This would represent the guess that M may already have been achieved. We shall, in fact, make this inclusion assumption; that is, <u>if</u> $B_{k+1}(x_1, x_2, \cdots, x_k)$ <u>is empty, then we take</u>

$$(2.9) \qquad F_{k+1}(x_1, x_2, \cdots, x_k) \supseteq A^*_{k+1}(x_1, x_2, \cdots, x_k) \quad .$$

We have now finished our specification of a feasible set of analytic conditions on the conditional probabilities (1.4), and we are ready to study the implications of these conditions. We direct our attention to obtaining a lower bound, in terms of the constants b_k and c_k, $k = 2, 3, \cdots, s$ for the comprehensive probability $P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\}$. Let us define another estimator $\left\{T_k^{\,o}, k = 1, 2, \cdots\right\}$ as follows:

$$(2.10) \qquad T_k^{\,o}(x_1, x_2, \cdots, x_k) = \max_{x=x_1,x_2,\cdots,x_k} f(x) \quad .$$

By virtue of the inclusion $\left\{x_1, x_2, \cdots, x_k\right\} \subseteq A_{k+1}(x_1, x_2, \cdots, x_k)$, we have always

$$(2.11) \qquad T_k^{\,o}(x_1, x_2, \cdots, x_k) \leq T_k(x_1, x_2, \cdots, x_k) \quad .$$

And therefore,

$$(2.12) \qquad \begin{aligned} &P\left\{T_s(X_1, X_2, \cdots, X_k) = M\right\} \geq \\ &P\left\{T_s^{\,o}(X_1, X_2, \cdots, X_s) = M\right\} , \quad s = 1, 2, \cdots \quad . \end{aligned}$$

Hence, we shall have a lower bound for the left-hand side of (2.12) when we have found a lower bound for the right-hand side.

Now, the right-hand side of (2.12) is equal to $1 - P\left\{T_s^{\circ}(X_1, X_2, \cdots, X_s) < M\right\}$ (cf. (1.1) and (2.11)), and therefore we may, and shall actually work first toward an upper bound for

$$(2.13)\quad P\left\{T_s^{\circ}(X_1, X_2, \cdots, X_s) < M\right\} = P\left\{f(X_i) < M,\ i = 1, 2, \cdots, s\right\} .$$

Toward this end we set down the following equalities and inequalities, derived from (2.1), (2.2), (2.6), and (2.9).

$$(2.14)\quad P\left\{f(X_{k+1}) < M \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} = 0 \ \underline{\text{if}}\ (1.7)\ \underline{\text{holds}} .$$

<u>If</u> (1.7) <u>does</u> <u>not</u> <u>hold,</u> <u>and</u> <u>if</u> $K \cap G_{k+1}(x_1, x_2, \cdots, x_k) \neq \Lambda$, <u>then</u>

$$(2.15)\quad P\left\{f(X_{k+1}) < M \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} \leq 1 - b_{k+1} .$$

<u>If</u> (1.7) <u>does</u> <u>not</u> <u>hold,</u> <u>and</u> <u>if</u> $K \cap G_{k+1}(x_1, x_2, \cdots, x_k) = \Lambda$, <u>then</u>,

$$(2.16)\quad P\left\{f(X_{k+1}) < M \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} \leq 1 - \nu\, c_{k+1} .$$

These three assertions, for $k = 2, 3, \cdots, s - 1$, together with

$$P\left\{f(X_1) < M\right\} \leq 1 ,$$

are now applied to the relations

$$
(2.17) \quad \begin{aligned} & P\left\{f(X_i) < M,\ i = 1, 2, \cdots, k+1\right\} = \\ & \sum_{\substack{f(x_j) < M \\ j=1,2,\cdots,k}} P\left\{f(X_{k+1}) < M \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\} \\ & \times P\left\{X_i = x_i,\ i = 1, 2, \cdots, k\right\},\ k = 2, 3, \cdots, s-1 . \end{aligned}
$$

And the result in general is the following:

$$
(2.18) \quad \begin{aligned} & P\left\{f(X_i) < M,\ i = 1, 2, \cdots, s\right\} \leq \\ & \prod_{j=2}^{s} (\max\,[1 - b_j,\ 1 - \nu c_j]),\ s = 2, 3, \cdots , \end{aligned}
$$

whence

$$
(2.19) \quad \begin{aligned} & P\left\{T_s^{\,o}(X_1, X_2, \cdots X_s) = M\right\} \geq \\ & 1 - \prod_{j=2}^{s} (\max\,[1 - b_j,\ 1 - \nu c_j]),\quad s = 2, 3, \cdots . \end{aligned}
$$

We have said that (2.18) is the consequence of our calculations <u>in general</u>. One finds, in carrying out those calculations, that the following hold:

<u>If for all</u> $k = 2, 3, \cdots$, <u>and all</u> $x_1, x_2, \cdots, x_k$, $G_{k+1}(x_1, x_2, \cdots, x_k)$ <u>contains a point of</u> K, <u>then</u>

$$
(2.20) \quad \begin{aligned} & P\left\{T_s^{\,o}(X_1, X_2, \cdots, X_s) = M\right\} \geq \\ & 1 - \prod_{j=2}^{s} (1 - b_j),\quad s = 2, 3, \cdots . \end{aligned}
$$

<u>If for all</u> $k = 2, 3, \cdots$, <u>and all</u> $x_1, x_2, \cdots, x_k$, $G_{k+1}(x_1, x_2, \cdots, x_k)$ <u>contains no point of</u> K, <u>then</u>

$$P\left\{T_s^{\,o}(X_1, X_2, \cdots, X_s) = M\right\} \geqq \tag{2.21}$$
$$1 - \prod_{j=2}^{s} (1 - \nu c_j), \quad s = 2, 3, \cdots .$$

The import of these special results is this. Suppose the computer has great confidence in his sets $G_{k+1}(x_1, x_2, \cdots, x_k)$, that is, confidence that they contain points of K. Then he will make the numbers b_j big (i.e., very close to 1) with the result that they will be much bigger than the corresponding numbers νc_j. Then if, although this is not within his information, the sets $G_{k+1}(x_1, x_2, \cdots, x_k)$ do in fact all contain points of K(i.e., his confidence in these sets is, unbeknown to him, actually verified), the probabilities $P\left\{T_s^{\,o}(X_1, X_2, \cdots, X_s) = M\right\}$ have the larger lower bounds $1 - \prod_{j=2}^{s} (1 - b_j)$. On the other hand, if, in fact, none of the sets G_{k+1} contains a point of K (in spite of his confidence) then we can only assert (2.21), with the smaller lower bounds $1 - \prod_{j=2}^{s} (1 - \nu c_j)$.

Now, a necessary and sufficient condition that the products in (2.18) and (2.19), for $s = 2, 3, \cdots$, tend to 0 is that the series

$$\sum_{j=2}^{\infty} \min\,[b_j, \nu c_j]$$

be divergent. This condition, which, through (2.19) and (2.12) is seen to be a sufficient condition for weak convergence of $\mathcal{J}$ to M, is

however, not a feasible condition, for, we do not know the value of ν. However, by virtue of (2.5) we have

$$\sum_{j=2}^{s} \min\,[b_j, \nu c_j] > \sum_{j=2}^{s} c_j, \quad s = 2, 3, \cdots .$$

Hence, we can take the feasible condition that $\Sigma\, c_j$ be divergent. This has, in addition, the consequence (again through (2.5)) that also $\Sigma\, b_j$ is divergent. Thus, the convergence to 1 of the right-hand sides of all three inequalities (2.19), (2.21), and (2.20) is assured.

Hence, the divergence of $\Sigma\, c_j$ insures the weak convergence of $\mathcal{T}$ to M in any case. But more than this, it insures the strong convergence of $\mathcal{T}$ to M. This is a consequence of (1.2). That is, by virtue of (1.2) we have, for each $s = 1, 2, \cdots$,

$$(2.22)\quad \begin{aligned} &\left\{(x_1, x_2, \cdots) \in D \times D \times \cdots \mid T_s(x_1, x_2, \cdots, x_s) = M\right\} \\ &\subseteq \left\{(x_1, x_2, \cdots) \in D \times D \times \cdots \mid \lim_{r\to\infty} T_r(x_1, x_2, \cdots, x_r) = M\right\} \end{aligned}$$

which implies that, for each $s = 1, 2, \cdots$,

$$(2.23)\quad \begin{aligned} &P\left\{\lim_{r\to\infty} T_r(X_1, X_2, \cdots, X_r) = M\right\} \geq \\ &P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\}, \end{aligned}$$

so that weak convergence implies strong convergence.

If we combine (2.12) with the inequalities (2.19), (2.20), and (2.21) to write: <u>in general</u>

$$P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\} \geq 1 - \prod_{j=2}^{s} (\max[1 - b_j, 1 - \nu c_j]), \quad s = 2, 3, \cdots ; \tag{2.24}$$

if for all $k = 2, 3, \cdots$, and all $x_1, x_2, \cdots, x_k$, $G_{k+1}(x_1, x_2, \cdots, x_k)$ contains a point of K, then

$$P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\} \geq 1 - \prod_{j=2}^{s} (1 - b_j), \quad s = 2, 3, \cdots ; \tag{2.25}$$

if for all $k = 2, 3, \cdots$, and all $x_1, x_2, \cdots, x_k$, $G_{k+1}(x_1, x_2, \cdots, x_k)$ contains no point of K, then

$$P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\} \geq 1 - \prod_{j=2}^{s} (1 - \nu c_j), \quad s = 2, 3, \cdots ; \tag{2.26}$$

then we can state our result as follows:

THEOREM 1. *If the stochastic process* $\mathcal{S}$ *satisfies the conditions* (2.1), (2.2), (2.5), (2.6), and (2.9), *then the inequalities* (2.24), (2.25), and (2.26) *hold in the respective cases described. And in any case a sufficient condition that* $\mathcal{T} = \left\{T_s(X_1, X_2, \cdots, X_s), s = 1, 2, \cdots\right\}$ *converge almost surely to* M *is that the series* $\Sigma\, c_j$ *be divergent.*

This theorem actually states a weaker convergence result than we have established. That is, (1.2) and consequently (2.22) and (2.23)

hold also with T^o in place of T. Therefore, through the inequalities (2.19), (2.20), and (2.21), the divergence of $\Sigma\, c_j$ implies almost sure convergence of $\mathcal{J}^o = \left\{T_s{}^o(X_1, X_2, \cdots, X_s), s = 1, 2, \cdots\right\}$ to M. And this is the stronger result by virtue of (2.11) and (1.1). However, our attention is concentrated on the random variables $T_s(X_1, X_2, \cdots, X_s)$, and this being the case, our <u>detailed</u> results, namely the inequalities (2.24), (2.25), and (2.26) are in fact the weaker for having come about by our taking advantage of (2.12). We shall say more about this in section 4.

3. NECESSARY CONDITIONS FOR STRONG CONVERGENCE OF $\mathcal{J}$ TO M

In this section we shall obtain upper bounds for the probabilities $P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\}$, $s = 1, 2, \cdots$, and from them obtain a necessary condition for strong convergence of $\mathcal{J}$ to M.

Let us define

(3.1)
$$A^o_{k+1}(x_1, x_2, \cdots, x_k) = \left\{x \in A_{k+1}(x_1, x_2, \cdots, x_k) \mid f(x) \text{ is known}\right\} ,$$

and

(3.2)
$$a_{k+1} = \min_{x_1, x_2, \cdots, x_k} P\left\{X_{k+1} \in A^o_{k+1}(x_1, x_2, \cdots, x_k) \mid X_i = x_i,\ i = 1, 2, \cdots, k\right\}$$

$$k = 1, 2, \cdots .$$

It is to be noted that the a_k are feasible parameters; that is, the computer can control them (to within the requisite guessing).

Now let us direct our attention to the following important subsets of D:

$$(3.3) \quad H_k(x_1, x_2, \cdots, x_{k-1}) = \left\{ x_k \in D \,|\, T_k(x_1, x_2, \cdots, x_k) < M \right\} .$$

For k = 1 we set also

$$(3.4) \qquad H_1 = \left\{ x \in D \,|\, T_1(x) < M \right\} .$$

By virtue of (1.2) we have

$$(3.5) \qquad \begin{cases} \underline{\text{if}}\ x_k \in H_k(x_1, x_2, \cdots, x_{k-1}) \ , \ \underline{\text{then}} \\ x_{k-1} \in H_{k-1}(x_1, x_2, \cdots, x_{k-2}) \ . \end{cases}$$

An equivalent statement is the following:

$$(3.6) \qquad \begin{cases} \underline{\text{if}}\ x_{k-1} \in \bar{H}_{k-1}(x_1, x_2, \cdots, x_{k-2}) \ , \ \underline{\text{then}} \\ H_k(x_1, x_2, \cdots, x_{k-1}) = \Lambda \ . \end{cases}$$

A consequence of this is that <u>if</u> $H_{k-1}(x_1, x_2, \cdots, x_{k-2}) = \Lambda$, <u>then</u> <u>for</u> <u>each</u> $x_{k-1} \in D$ <u>also</u> $H_k(x_1, x_2, \cdots, x_{k-1}) = \Lambda$.

From (3.5) or (3.6) we get the following identity:

(3.7) $$P\left\{T_s(X_1, X_2, \cdots, X_s) < M\right\} =$$

$$\sum_{\substack{x_1 \varepsilon H_1 \\ x_2 \varepsilon H_2(x_1) \cdots \\ x_{s-1} \varepsilon H_{s-1}(x_1, x_2, \cdots, x_{s-2})}} P\left\{X_s \varepsilon H_s(x_1, x_2, \cdots, x_{s-1}) \,|\, X_i = x_i,\ i = 1,2,\cdots, s-1\right\} \times$$

$$P\left\{X_{s-1} = x_{s-1} | X_i = x_i,\ i = 1,2,\cdots, s-2\right\} \times \cdots$$

$$P\left\{X_2 = x_2 \mid X_1 = x_1\right\} \times$$

$$P\left\{X_1 = x_1\right\} .$$

The importance of the sets H_k is that they provide this exact expression for $P\left\{T_s(X_1, X_2, \cdots, X_s) < M\right\}$.

If it were the case that $T_s(x_1, x_2, \cdots, x_s) \equiv M$ for all s and all $x_1, x_2, \cdots, x_s$, then (3.7) would vanish identically. However, we are excluding this extraordinary situation, and therefore we are assured that for each k there is at least one set $H_k(x_1, x_2, \cdots, x_{k-1})$ which is not empty.

Now let us take note of our information scheme. If, at the (k-1)th stage, the knowledge of the values of f on $A_k^{\,o}(x_1, x_2, \cdots, x_{k-1})$ could in itself improve upon the number $T_{k-1}(x_1, x_2, \cdots, x_{k-1})$, then in the information at that stage we should have put, not $T_{k-1}(x_1, x_2, \cdots, x_{k-1})$, but the improved number. In other words, it is implicit in the information scheme that if the k-th realized point, x_k, is, as a point of D, a duplicate of one of the points in $A_k^{\,o}(x_1, x_2, \cdots, x_{k-1})$ then $T_k(x_1, x_2, \cdots, x_k) = T_{k-1}(x_1, x_2, \cdots, x_{k-1})$ -- that is, the T-value

does not increase. In particular, therefore, if $T_{k-1}(x_1, x_2, \cdots, x_{k-1}) < M$, then for each $x \in A_k{}^{o}(x_1, x_2, \cdots, x_{k-1})$ we have $T_k(x_1, x_2, \cdots, x_{k-1}, x) < M$. Otherwise expressed:

$$(3.8) \qquad \begin{cases} \underline{\text{if}}\ x_{k-1} \in H_{k-1}(x_1, x_2, \cdots, x_{k-2}) \ , \ \underline{\text{then}} \\ A_k{}^{o}(x_1, x_2, \cdots, x_{k-1}) \subseteq H_k(x_1, x_2, \cdots, x_{k-1}) \ . \end{cases}$$

Combining this with (3.6) we get:

$$(3.9) \qquad \begin{cases} \underline{\text{if}}\ H_k(x_1, x_2, \cdots, x_{k-1}) \neq \Lambda \ , \ \underline{\text{then}} \\ A_k{}^{o}(x_1, x_2, \cdots, x_{k-1}) \subseteq H_k(x_1, x_2, \cdots, x_{k-1}) \ . \end{cases}$$

From (3.2) and (3.9) we get the following inequalities:

$$(3.10) \qquad P\left\{X_k \in H_k(x_1, x_2, \cdots, x_{k-1}) \,|\, X_i = x_i,\ i = 1, 2, \cdots, k-1\right\} \geq a_k \ ,$$
$$k = 2, 3, \cdots \ .$$

If, now, we set

$$P\left\{X_1 \in H_1\right\} = a_1 \ ,$$

and apply (3.10) to (3.7), we obtain the following result:

$$(3.11) \qquad P\left\{T_s(X_1, X_2, \cdots, X_s) < M\right\} \geq \prod_{j=1}^{s} a_j \ .$$

And therefore,

$$(3.12) \qquad P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\} \leq 1 - \prod_{j=1}^{s} a_j \ .$$

If there is strong convergence of $\mathcal{T}$ to M, then the left-hand side of (3.12) must tend to 1 as $s \to \infty$. Consequently we must have $\prod_{j=2}^{s} a_j \to 0$ (by our assumption of a non-trivial problem, $a_1 > 0$). We state this result in

THEOREM 2. <u>A necessary condition that the stochastic process give strong (or weak, since the two are equivalent by virtue of (1.2) convergence of $\mathcal{T}$ to M is that</u> (cf. (3.2))

$$(3.13) \qquad \lim_{s \to \infty} \prod_{j=2}^{s} a_j = 0 \quad .$$

<u>The inequalities (3.12) hold for $\mathcal{S}$.</u>

It is readily seen that the relations

$$(3.14) \qquad a_i \leqq 1 - c_i \quad , \quad i = 2, 3, \cdots$$

hold, so that the divergence of $\Sigma\, c_i$ implies (3.13).

This theorem is in the nature of an exact verification of the first motivation discussed in section 2. It says, roughly, that we must not consistently allot too much conditional probability to the points at which the value of f is already known, if we are to have stochastic convergence of $\mathcal{T}$ to M.

By virtue of (2.11), the inequalities (3.12) and the theorem hold a fortiori if T is replaced by T°.

4. CONCERNING SHARPENING OF THE RESULTS

The results we have obtained in section 2 are quite obviously obtained actually for the estimator $\left\{T_s^o, s = 1, 2, \cdots\right\}$, and come to bear upon $\left\{T_s, s = 1, 2, \cdots\right\}$ through (2.12). This means that those results could have been obtained even if we had not introduced the hypothesis of additional k-th stage information embodied in the generally conceived sets $A_{k+1}(x_1, x_2, \cdots, x_k)$ and $B_{k+1}(x_1, x_2, \cdots, x_k)$; that is, if we had assumed $\mathcal{J}_k^o(x_1, x_2, \cdots, x_k)$ as the only k-th stage information. And the results in section 3 obtain by the grace of (3.9), which is a completely general relation, therefore holding also in the special case where $\mathcal{J}_k^o(x_1, x_2, \cdots, x_k)$ is the only information.

One of the directions for investigation after bounds which will be improvements on (2.24), (2.25), (2.26), and (3.12), is in making additional assumptions concerning the information scheme. For example, if we specialize to the situation that D is a rectangular lattice in the plane,

$$D = \left\{ (mh, nh) \mid m, n = 1, 2, \cdots, r \right\},$$

and that the computer <u>knows</u> f to have the property

$$|f(mh, nh) - f(m'h, n'h)| < \alpha$$

$$\underline{\text{for}}\ m' = m \pm 1, \quad n' = n \pm 1 \quad ,$$

where α is a fixed number, then the following can be said: if $x_j(j = 1, 2, \cdots, k)$ is such that $f(x_j) + \alpha \leq \max_{i+1,2,\cdots k} f(x_i)$,

then the four points $x_j + (h, h)$, $x_j + (-h, h)$, $x_j + (h, -h)$, and $x_j + (-h, -h)$ all belong to $A_{k+1}(x_1, x_2, \cdots, x_k)$.

Such additional facts concerning the sets A_{k+1}, B_{k+1}, and F_{k+1} would enable us in general to improve the inclusion relation (3.9); that is, enable us to assert a larger set than $A_k^{\,o}$ included in H_k. In this way we could obtain smaller upper bounds for $P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\}$ than given in (3.12).

It is easy to see that we always have:

$$H_k(x_1, x_2, \cdots, x_{k-1}) \subseteq \bar{K} \quad . \tag{4.1}$$

This relation can be used with (3.7) to obtain lower bounds for the probabilities $P\left\{T_s(X_1, X_2, \cdots, X_s) = M\right\}$. <u>And these lower bounds are precisely</u> (2.24), (2.25), <u>and</u> (2.26). Now, additional facts concerning the sets A_{k+1}, B_{k+1}, and F_{k+1} would in general enable us also to improve (4.1), and thereby, through (3.7) to obtain better bounds than given in (2.24), (2.25), and (2.6).

Of course, such improved bounds would in turn give us less stringent sufficient conditions for convergence of $\mathcal{J}$ to M and more complete necessary conditions.

But even without making additional assumptions concerning the information scheme, there is a refinement that can be made to improve our results. This is to introduce more control parameters than the c_k, b_k, and a_k; namely, different parameters for different

values of $\rho(x_1, x_2, \cdots, x_k)$, $\tau(x_1, x_2, \cdots, x_k)$, and $\lambda(x_k, x_2, \cdots, x_k)$ ($\equiv$ number of points in $A^0_{k+1}(x_1, x_2, \cdots, x_k)$) as well as for different stages k. This would call for more refined calculations with (2.17) and (3.7).

August 12, 1953

THE NATIONAL BUREAU OF STANDARDS

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended by Congress in Public Law 619, 1950. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to Government Agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services and various consultation and information services. A major portion of the Bureau's work is performed for other Government Agencies, particularly the Department of Defense and the Atomic Energy Commission. The scope of activities is suggested by the listing of divisions and sections on the inside of the front cover.

Reports and Publications

The results of the Bureau's work take the form of either actual equipment and devices or published papers and reports. Reports are issued to the sponsoring agency of a particular project or program. Published papers appear either in the Bureau's own series of publications or in the journals of professional and scientific societies. The Bureau itself publishes three monthly periodicals, available from the Government Printing Office: The Journal of Research, which presents complete papers reporting technical investigations; the Technical News Bulletin, which presents summary and preliminary reports on work in progress; and Basic Radio Propagation Predictions, which provides data for determining the best frequencies to use for radio communications throughout the world. There are also five series of nonperiodical publications: The Applied Mathematics Series, Circulars, Handbooks, Building Materials and Structures Reports, and Miscellaneous Publications.

Information on the Bureau's publications can be found in NBS Circular 460, Publications of the National Bureau of Standards ($1.00). Information on calibration services and fees can be found in NBS Circular 483, Testing by the National Bureau of Standards (25 cents). Both are available from the Government Printing Office. Inquiries regarding the Bureau's reports and publications should be addressed to the Office of Scientific Publications, National Bureau of Standards, Washington 25, D. C.

NBS